DOCUMENT RESUME

ED 464 928 TM 033 858

AUTHOR Frye, Ann W.; Szauter, Karen M.; Litwins, Stephanie D.; Ofoegbu, Esther N.

TITLE Validating Processes for Using Curriculum Objectives as Standards in Curriculum Evaluation.

PUB DATE 2002-04-03

NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *College Faculty; *Curriculum; *Data Collection; Educational Objectives; *Generalizability Theory; Higher Education; Medical Education; *Medical Students; *Validity

IDENTIFIERS University of Texas Medical Branch Galveston

ABSTRACT

This paper describes the processes for collecting information about the objectives of a medical school curriculum and the validation studies of those processes. Major revisions were implemented in the curriculum of the first 2 years at the University of Texas Medical Branch (UTMB) in 1998. In the 2000-2001 academic year, student and faculty groups were asked to reflect on the curriculum objectives using processes tailored to each group's perspective. First and second year medical students (42 to 102 raters per course) documented the degree of emphasis on each objective they experienced in each course. Graduating students (n=170) recorded the emphasis on each objective experienced in the third and fourth years. Curriculum Committee members (n=15) recorded the degree of emphasis they thought should be given to each objective across the 4 years. The framework of generalizability theory allowed the evaluation of validity of these approaches. The design of the data collection processes and the results of generalizability analyses provide good evidence of the validity of the mean emphasis ratings generated by the processes described in this paper. An appendix contains an example of the survey questions and responses. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 464 928

Validating Processes for Using Curriculum Objectives as Standards in Curriculum Evaluation

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

A. Frye

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Ann W. Frye, Karen M. Szauter, Stephanie D. Litwins
The University of Texas Medical Branch
and
Esther N. Ofoegbu
The University of Nigeria Teaching Hospital

Presented at the 2002 Annual Meeting of the American Educational Research Association
New Orleans

April 3, 2002

TM033858

Contact: Ann W. Frye, Office of Educational Development, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555-0664 (awfrye@utmb.edu)

BEST COPY AVAILABLE

Validating Processes for Using Curriculum Objectives as Standards in Curriculum Evaluation

Ann W. Frye[1], Karen M. Szauter[1], Stephanie D. Litwins[1], Esther N. Ofoegbu[2]

[1]The University of Texas Medical Branch, Galveston, Texas
[2] The University of Nigeria Teaching Hospital, Enugu, Nigeria

Medical schools around the world are adopting new approaches to education. The success of these changes may depend on careful monitoring, particularly when the changes challenge a school's longstanding practices. Periodic assessment against dependable standards has been found useful for maintaining the outcomes of radical change (Gerrity and Mahaffy, 1998; Robins, White, and Fantone, 2000). In particular, when changes apply to the entire curriculum, periodic measurement against curriculum objectives can determine if the curriculum remains aligned with those objectives. Information about a curriculum's match to its objectives is potentially available from the students who experience the curriculum, the faculty members who plan or lead its instructional units, and the faculty policymakers responsible for the entire curriculum. Analysis of information from these groups may be useful for defining curricular strengths and weaknesses. This paper describes the processes for collecting information about a curriculum's objectives and the validation studies of those processes.

The medical school faculty of The University of Texas Medical Branch (UTMB) approved 29 curriculum-level objectives to guide the revision of its four-year medical curriculum. The objectives were adapted from the Medical School Objectives Project (MSOP) presented in a report prepared by the Association of American Medical Colleges' Medical School Objectives Writing Group (1998) and published for a wider audience in 1999 (Medical School Objectives Writing Group, 1999). The MSOP objectives were developed to guide medical schools in shaping their curricula so that medical school graduates were prepared for the demands of post-graduate training and contemporary medical practice. UTMB's adaptation of the MSOP objectives (see Appendix I) outlined the objectives to be addressed by the sum total of all courses, clerkships, and other learning opportunities composing the four-year curriculum. In concert with the language of the MSOP report and UTMB curriculum documents, we will use the term "curriculum objectives" in this paper, recognizing that some educators might contend that these "objectives" are not framed in the traditional specific-and-measurable language recommended for writing learning objectives.

Major revisions to the first two years of the UTMB medical school curriculum, traditionally known as the "basic science years", were implemented in the fall of 1998. The clinical clerkships and courses of the curriculum's third and fourth year are presently under review. The changes in the first and second year curriculum were radical, involving reorganization of content and the introduction of new teaching methods (Bernier, Adler, Kanter and Meyer, 2000). Discipline-based lecture courses (e.g., Biochemistry, Physiology) were replaced with

interdisciplinary courses organized around human organ systems, featuring problem-based learning. The new courses integrated basic and clinical science concepts and were complemented by courses introducing students to the practice of medicine.

Recognizing the need for reliable information to inform continued curriculum evolution and to detect unwanted "drift" back to the familiar old approaches, (Robins, White, and Fantone, 2000), the Dean convened a faculty working group to design and implement a comprehensive curriculum evaluation plan. One element of the plan proposed using the 29 curriculum objectives as standards against which to measure the curriculum. We expected that each objective would be addressed several times in the curriculum. For example, the objective "knowledge of the normal structure and function of the body and of each of its major organ systems" was likely to be a part of most, perhaps all, courses to some degree. We anticipated, however, that some objectives might be less adequately addressed in the new curriculum that depended on newly developed inter-departmental and cross-disciplinary structures for adjusting itself. For example, the objective "a commitment to provide care to patients who are unable to pay and to advocate for access to health care for members of traditionally underserved populations" could potentially be addressed by many courses but also might easily be overlooked in the reorganization and subsequent fine-tuning of the curriculum. Similarly, the inter-disciplinary nature of the new courses increased the risk that unrecognized redundancy among courses would result in over-emphasis on other objectives. We therefore developed processes to determine in which instructional units curriculum objectives were being emphasized. We wanted data that would allow us to judge whether each objective was receiving <u>enough</u> emphasis in the curriculum. We also wished to examine any differences between students' experiences with the objectives and faculty intentions for the curriculum. Sound data from these processes could be used to inform decisions about adding or subtracting topics and emphases from individual courses. Throughout the remainder of the paper, we will use the term "curriculum unit" to refer to the discrete instructional units (e.g., courses, clerkships) as well as to the collections of instructional units (e.g., all third and fourth year required clerkships) involved in the studies.

In the 2000-2001 academic year, student and faculty groups were asked to reflect on the curriculum objectives using processes tailored to each group's perspective. We planned to compile these data to reflect on the match of the curriculum to the curriculum objectives. After each first and second year course, students documented the degree of emphasis on each objective they *experienced* in that course. Graduating students recorded the degree of emphasis on each objective *experienced* across the third and fourth years (the clinical clerkships) as a unit. Course and clerkship directors recorded the degree of emphasis they *intended* for each objective in their curricular unit. Curriculum Committee (CC) members, the curriculum policymaking group, recorded the degree of emphasis they thought *should be given* to each objective across Years 1 and 2 and across Years 3 and 4. Mean "emphasis" ratings for each objective could then be constructed for use in assessments of the curriculum against the

objectives. Differences in responses of students, faculty, and policy setters could also be investigated using the mean ratings. Because the data collection processes were new to the school, careful investigations into validity were conducted before using the data in curriculum evaluation procedures.

Theoretical Framework for Validity Studies

We used several methods to investigate the validity of the measures and the score data collected from them. We studied the design of the data collection processes, adjusted the wording of the items to fit each group, and computed appropriate generalizability coefficients for the scales.

Careful design of the data collection processes allayed some validity concerns. For example, all groups with an investment in the curriculum (students at all levels, course and clerkship directors, faculty policy setters) were asked questions about the objectives that their experience or position prepared them to answer.

We addressed other validity concerns through an examination of data collected from the various groups by the prescribed processes. The framework of generalizability theory allowed us to assess the effect of some potential threats to validity. Generalizability theory supports inquiries into the proportion of variance in a data set attributable to "true score" and to error variance. Reliability coefficients can be constructed to summarize the relative contributions of true-score and error variance.

Two validity-related premises were investigated using generalizability theory. First, for a given group, valid mean ratings should demonstrate differentiation between the objectives; that is, raters would not give every objective similar emphasis ratings, leading to similar mean ratings for all objectives. For example, in the second-year Cardiovascular/Pulmonary course, students' mean ratings of the emphasis given to objectives such as "*knowledge of the normal structure and function of the body and of each of its major organ systems*" would be expected to differ from the mean rating of the course's emphasis on objectives such as "*knowledge of various approaches to the organization, financing, and delivery of health care*". If a group's mean ratings for objectives within a given course or clerkship were all similar, then the validity and usefulness of the ratings would be compromised.

Secondly, evidence that raters were in relative agreement on the rating for each objective in a given curriculum unit would be an important indicator of validity. Some variability among raters was expected, but the degree of rater agreement would provide strong evidence of whether or not the mean objective ratings could fairly represent the views of the rater group.

In the generalizability analyses described in this paper, differences between objectives' ratings are assumed to reflect valid variability. Differences between raters and other sources of rating variability are defined as error. In each study, *raters* are treated as a sample from the universe of possible raters of that type (e.g., first-year students) and *objectives* as a sample from the universe of possible curriculum objectives, both random facets. The curricular unit (course

or 2-year span) is treated as a fixed facet, with data analyzed separately for each (Shavelson and Webb, 1991).

We did not apply generalizability analyses to rating data obtained from the course and clerkship directors. In most cases, courses and clerkships were each rated by a single faculty rater. Those studies are therefore not represented in this discussion.

We used two statistics to summarize the validity evidence for the mean emphasis ratings collected in processes employing more than one rater per curricular unit. The *phi* coefficient, also known as the *dependability* coefficient, is the most germane reliability coefficient available from generalizability analyses (Shavelson and Webb, 1991). Because it estimates the reliability (reproducibility or dependability) of the value of each objective's mean rating, it is more informative for our work than the more familiar G coefficient, which estimates reliability of ratings' rank orders. Our planned use of the ratings data required dependable mean emphasis ratings for each objective rather than a dependable rank ordering of the objectives' ratings. Using generalizability theory's variance component estimates, the *phi* coefficient contrasts "true" or desirable variance in objectives' ratings to the amount of undesirable variability among raters on each objective plus additional error variance. We arbitrarily defined a *phi* of .8 or greater as sufficient evidence for validity. The size of *phi* for any data set is affected by the relative size of variance components associated with the study design (*objectives, raters, error*) and by the number of raters. A larger proportion of variance attributable to differences in objectives' mean ratings ("true" variance) and larger numbers of raters (which decreases mean variance due to rater variability) are both expected to be associated with higher *phi* coefficients.

The second statistic used to describe validity of the mean ratings was the range of high and low values of the mean emphasis ratings in each data set. The range indicates the degree of discrimination between objectives achieved by raters in that study.

Methods

The methods for the three data collection processes subjected to generalizability analysis are each described separately. We describe the elements common to all studies first.

Common Elements

In each study, raters considered the emphasis given to each objective in the context of the specified course or two-year span of courses. The raters were students in Studies I and II and faculty in Study III. A Web-based or paper questionnaire presented the 29 objectives grouped under four basic attributes of physicians ("knowledgeable", "skillful", "professional", "life-long learners"). The reader is referred to Appendix I for an illustration of pertinent elements of the questionnaire. All studies employed the same 6-point response scale, ranging from 0="*imperceptible (no) emphasis*" to 5="*heavy emphasis*". Raters were instructed that objectives were expected to have different degrees of emphasis and that some objectives might receive no emphasis at all in a particular unit.

Regardless of presentation format, the objectives emphasis questionnaire took approximately 10 minutes to complete.

Study I: First and second year courses rated by students

First and second year medical students completed confidential Web-based objectives emphasis questionnaires at the end of each of the 14 Year 1 and 2 courses in academic year 2000-2001. The first and second year classes were each composed of approximately 200 students throughout the year. At the end of each course, we randomly assigned students to respond to either the objectives questionnaire or other course-related surveys. As a result, either ¼ or ½ of the appropriate class completed the objectives questionnaire for each course, with the number of respondents dependent on how many different surveys were distributed at that time. The number of raters ranged from 42 to 102 per course in Study I.

Responses for each course were analyzed separately, since each was evaluated by a different group of students. Each course's data were fit to an *objective+rater+error=rating* model in which *objective* and *rater* were random facets.

Study II: Third and fourth year required clerkships rated by students

Graduating medical students (n=170) in an end-of-year meeting recorded the degree of emphasis given to each objective across the combined third and fourth (clinical) years. They used a paper version of the objectives emphasis questionnaire. These response data were also fit to an *objective+rater+error=rating* model in which *objective* and *rater* were random facets.

Study III: Years 1&2 and years 3&4 rated by faculty policy setters

Fifteen of the 16 Curriculum Committee (CC) members each recorded the degree of emphasis that they thought should be given to each objective in the first two years and in the last two years of the curriculum. We modified the layout of the paper questionnaire to accommodate side by side responses for both curricular units. The two sets of ratings in Study III were separately fit to *objective+rater+error=rating* models.

Results

Phi coefficients of .83 or greater were obtained for ratings from all courses in Study I and for the ratings in Studies II and III. In Studies I and II, larger numbers of raters tended to be associated with larger phi coefficient values. The lowest phi coefficient observed in these studies, .83, was associated with the smallest range of mean ratings observed for a course, 2.3 to 3.4, all moderate ratings. For an illustration of the ratings for a single curriculum unit, the reader is referred to Appendix I.

In Study III, CC members rated the ideal level of emphasis for objectives in both Years 1&2 and Years 3&4. Phi values of .92 for Years 1&2 and .83 for Years 3&4 were obtained, both of acceptable magnitude. The .92 phi coefficient

value was associated with a wider range of mean emphasis ratings, indicating that the raters recorded greater differences between objectives in Years 1&2 than in Years 3&4. Mean ratings for Years 1&2 ranged from 1.5 to 4.9 (low to very high), as contrasted to mean ratings for Years 3&4, which ranged from 3.2 to 5.0 (moderate to very high). Table I summarizes the validity evidence accumulated in Studies I-III.

Table I: Validity Evidence Summary

Study	Unit rated	Rater	Number of raters	Phi coefficient	Span of mean ratings
I	Individual courses in Years 1&2	Students	42-102	From .83-.97 across all 14 courses	Largest span for a course: 1.2 to 4.6 Smallest span for a course: 2.3 to 3.4
II	Years 3&4 as unit	Students	170	.97	2.4 to 4.4
III	Years 1&2 as unit	CC members	15	.92	1.5 to 4.9
III	Years 3&4 as unit	CC members	15	.83	3.2 to 5.0

Discussion

The design of the data collection processes (asking appropriate groups of raters questions that they should be able to answer based on their experience or positions) along with the results of generalizability analyses provide good evidence for validity of the mean emphasis ratings generated by the processes described in this paper. All phi coefficients for Studies I-III exceeded .8, indicating adequate differentiation among objectives and adequate agreement among raters. The mean ratings for objectives within each study were sufficiently different that meaningful comparisons of the ratings could be made. The range

of mean ratings in each study provides some direct description of that feature. The degree of agreement among groups of raters was strong enough to support the use of mean ratings in analyses.

Although the presence of these good qualities is not proof of the validity of the mean emphasis ratings, their absence would certainly have indicated problems. It would be difficult, for example, to trust the use of a mean rating in an analysis if there were substantial disagreement among the raters' responses. Similarly, it would be hard to contrast mean ratings if they all had similar values, regardless of differences in the likelihood that all objectives were equally well represented in the curriculum unit. These studies' results cumulatively suggest that mean ratings of emphasis on curriculum objectives in different curriculum units obtained by these processes can be used with confidence in subsequent curriculum evaluation studies.

Curriculum objectives are well understood as guides for developing curriculum outcome measures. Their use as standards against which a curriculum may be assessed is less well explored. The studies described in this paper suggest that both students and faculty provided believable ratings of emphasis on curriculum objectives from their own perspectives. The ratings should be useful to examine important aspects of the curriculum. Similar processes may be useful to any professional curriculum undergoing significant change.

References

Association of American Medical Colleges (1998). Report I: Learning Objectives for Medical Student Education – Guidelines for Medical Schools. Available online from http://www.aamc.org/meded/msop/start.htm.

Bernier, G.M., Adler, S., Kanter, S., & Meyer, W.J. (2000). On changing curricula: lessons learned at two dissimilar medical schools. Academic Medicine, 75(6), 595-601.

Gerrity, M. S. & Mahaffy, J. (1998). Evaluating change in medical school curricula: How did we know where we were going? Academic Medicine, 73 (9 suppl), s55- s59.

Robins, L.S., White, C.B. & Fantone J.C. (2000). The difficulty of sustaining curricular reforms: A study of "drift" at one school. Academic Medicine, 75 (8), 801-805.

Shavelson, R.J. & Webb, N.M. (1991). Generalizability theory: A primer (Volume I). Newbury Park: Sage Publications.

The Medical School Objectives Writing Group (1999). Learning objectives for medical student education -- guidelines for medical schools: Report I of the Medical School Objectives Project. Academic Medicine, 74 (1), 13-18.

Appendix I

Example of Survey Instrument and Emphasis Ratings for a Second-Year Course

In this representation of the student questionnaire, the column in which respondents would have indicated their ratings has been used to list the mean student response on a 0-4 scale for each objective for a selected second-year course. Forty-five students responded to the questionnaire for this course.

Instructions
Please think about HOW MUCH EMPHASIS was placed on each of the following objectives by the course indicated above. Indicate your opinion by placing an "X" in the box under a number from 0 to 5 for each objective. Use "0" to indicate "an imperceptible (no) emphasis" and "5" to indicate "heavy emphasis". Use the numbers from 1 – 4 to indicate levels of more moderate emphasis.

A course may give "no" emphasis to any number of objectives, "moderate" emphasis to any number of objectives, and "heavy" emphasis to any number of objectives. Since these objectives are intended to cover all four years of medical school, no one course is likely to include an emphasis on all objectives.

How much emphasis was placed on each of these 11 objectives related to the overall curriculum goal "To produce knowledgeable physicians"?

	Mean rating
Knowledge of the theories and principles that govern **ethical decision making**, and of the major **ethical dilemmas** in medicine, particularly those that arise at the beginning and end of life, those that arise from the knowledge of genetics, and those that threaten medical professionalism posed by conflicts of interests inherent in various financial and organizational arrangements for the practice of medicine.	2.38
Knowledge of the **normal structure and function** of the body and of each of its major organ systems.	4.11
Knowledge of the **molecular, biochemical, and cellular mechanisms** that are important in maintaining the body's homeostasis.	4.04
Knowledge of the various **causes** (genetic, developmental, metabolic, toxic, microbiologic, autoimmune, neoplastic, degenerative, and traumatic) **of maladies** and the ways in which they operate in the body (**pathogenesis**).	3.73
Knowledge of the **altered structure and function (pathology and pathophysiology)** of the body and its major organ systems that are seen in various diseases and conditions.	4.07
Knowledge of the most frequent **clinical, laboratory, roentgenologic, and pathologic manifestations** of common maladies.	3.67
Knowledge about relieving **pain and ameliorating the suffering** of patients.	2.38
Knowledge of the **important non-biological determinants of poor health** and of the **economic, psychological, social, and cultural factors** that contribute to the development and/or continuation of maladies.	2.20
Knowledge of the **epidemiology** of common maladies within a defined population and the systematic approaches useful in reducing the incidence and prevalence of those maladies.	3.02
Knowledge of various approaches to the **organization, financing, and delivery of health care.**	1.36
An understanding of the **power of the scientific method** in establishing the causation of disease and efficacy of traditional and non-traditional therapies.	2.36

How much emphasis was placed on each of these 10 objectives related to the overall curriculum goal "To produce skillful physicians"?

	Mean rating
The ability to obtain an **accurate medical history** that covers all essential aspects of the history, including issues related to age, gender, and socio-economic status.	3.00
The ability to perform both a **complete and a focused examination**, including a mental status examination.	2.51
The ability to perform **routine technical procedures** including at a minimum venipuncture, inserting an intravenous catheter, arterial puncture, thoracentesis, lumbar puncture, inserting a nasogastric tube, inserting a foley catheter, and suturing lacerations.	1.67
The ability to interpret the results of commonly used **diagnostic procedures**.	3.24
The ability to **reason deductively** in solving clinical problems.	3.24
The ability to construct appropriate **management strategies** (both diagnostic and therapeutic) for patients with common conditions, both acute and chronic, including medical, psychiatric, and surgical conditions, and those requiring short-and long-term rehabilitation.	2.82
The ability to **recognize patients with immediately life threatening** *cardiac, pulmonary, or neurological* **conditions** regardless of etiology, and to **institute appropriate initial therapy**.	2.84
The ability to recognize and outline an **initial course of management** for patients with serious conditions requiring critical care.	2.80
The ability to **communicate effectively, both orally and in writing,** with patients, patients' families, colleagues, and others with whom physicians must exchange information in carrying out their responsibilities.	2.13
The ability to **identify factors that place individuals at risk** for disease or injury, to select appropriate **tests for detecting** patients at risk for specific diseases or in the early stage of disease, and to determine **strategies for responding** appropriately.	3.27

How much emphasis was placed on each of these 5 objectives related to the overall curriculum goal "To produce physicians possessing professional attitudes"?

	Mean Rating
An understanding of, and respect for, the **roles of other health care professionals**, and of the need to **collaborate** with others in caring for individual patients and in promoting the health of defined populations.	1.78
Compassionate treatment of patients, and **respect** for their privacy and dignity.	2.16
Honesty and integrity in all interactions with patient's families, colleagues, and others with whom physicians must interact in their professional lives.	2.27
A commitment to **advocate at all times the interests of one's patients** over one's own interest.	2.22
A commitment to **provide care to patients who are unable to pay** and to advocate for **access to health care** for members of traditionally underserved populations.	1.44

BEST COPY AVAILABLE

How much emphasis was placed on each of these 3 objectives related to the overall curriculum goal "To produce physicians committed to life-long learning"?

	Mean Rating
The capacity to **recognize and accept limitations** in one's knowledge and clinical skills, and a commitment to continuously improve one's knowledge and ability.	2.53
An understanding of the need to **engage in life-long learning** to stay abreast of relevant scientific advances, especially in the disciplines of genetics and molecular biology.	3.07
The ability to **retrieve** (from electronic databases and other resources), **manage**, and **utilize** biomedical information for solving problems and making decisions that are relevant to the care of individuals and populations.	2.87